

學號：B06942098 系級： 電信碩一 姓名：曾柏偉

1.請比較你實作的 generative model、logistic regression 的準確率，何者較佳？

答：logistic regression model 比較好，和老師上課說的因為 generative model 有個假設的關係。

	Generative model	Logistic model
accuracy	0.8451	0.8543

2.請說明你實作的 best model，其訓練方式和準確率為何？

答：

使用套件 XGBOOST，fit data（含相關係數較高的 feature），

準確率為 0.878000

3.請實作輸入特徵標準化(feature normalization)，並討論其對於你的模型準確率的影響。

答：generative model : 0.843925 → 0.84386 並無太大差異

logistic regression : 0.7812 → 0.85258 應該是 overflow 的關係，加上 feature 的分佈太散了。

4. 請實作 logistic regression 的正規化(regularization)，並討論其對於你的模型準確率的影響。

答：準確率幾乎是相等的，沒有很大的影響。大概都在 0.85257

5.請討論你認為哪個 attribute 對結果影響最大？

我覺得 education 及 occupation 的相關係數較大的種類數較多，但是 education 又更多了，所以我認為是 education。

	Bachelors	Doctorate	Master	Prof-school
correlation	0.18048	0.13257	0.17456	0.154906